Freeform Search

Database:
US Pre-Grant Publication Full-Text Database
US Patents Full-Text Database
US OCR Full-Text Database
EPO Abstracts Database
JPO Abstracts Database
Derwent World Patents Index
IBM Technical Disclosure Bulletins

Term: L4 and node$

Display: 100 **Documents in Display Format:** - **Starting with Number** 1

Generate: ○ Hit List ◉ Hit Count ○ Side by Side ○ Image

Search Clear Interrupt

Search History

DATE: Wednesday, November 09, 2005 Printable Copy Create Case

Set Name side by side	**Query**	**Hit Count**	**Set Name** result set
	DB=PGPB,USPT,USOC,EPAB,JPAB,DWPI,TDBD; PLUR=YES; OP=OR		
L6	L4 and node$	11	L6
L5	L4 and ((partition$ ordivid$) near node$)	1	L5
L4	L2 and (weight$ or rat$)	26	L4
L3	L2 and (weight$ near level)	1	L3
L2	L1 and ((backward or forward) near link$)	49	L2
L1	link$ near database	5261	L1

END OF SEARCH HISTORY

First Hit Fwd Refs Previous Doc Next Doc Go to Doc#

Generate Collection Print

L6: Entry 9 of 11 File: USPT Sep 4, 2001

DOCUMENT-IDENTIFIER: US 6285999 B1

TITLE: Method for node ranking in a linked database

Abstract Text (1):
A method assigns importance ranks to nodes in a linked database, such as any database of documents containing citations, the world wide web or any other hypermedia database. The rank assigned to a document is calculated from the ranks of documents citing it. In addition, the rank of a document is calculated from a constant representing the probability that a browser through the database will randomly jump to the document. The method is particularly useful in enhancing the performance of search engine results for hypermedia databases, such as the world wide web, whose documents have a large variation in quality.

Brief Summary Text (2):
This invention relates generally to techniques for analyzing linked databases. More particularly, it relates to methods for assigning ranks to nodes in a linked database, such as any database of documents containing citations, the world wide web or any other hypermedia database.

Brief Summary Text (11):
Various aspects of the present invention provide systems and methods for ranking documents in a linked database. One aspect provides an objective ranking based on the relationship between documents. Another aspect of the invention is directed to a technique for ranking documents within a database whose content has a large variation in quality and importance. Another aspect of the present invention is to provide a document ranking method that is scalable and can be applied to extremely large databases such as the world wide web. Additional aspects of the invention will become apparent in view of the following description and associated figures.

Brief Summary Text (12):
One aspect of the present invention is directed to taking advantage of the linked structure of a database to assign a rank to each document in the database, where the document rank is a measure of the importance of a document. Rather than determining relevance only from the intrinsic content of a document, or from the anchor text of backlinks to the document, a method consistent with the invention determines importance from the extrinsic relationships between documents. Intuitively, a document should be important (regardless of its content) if it is highly cited by other documents. Not all citations, however, are necessarily of equal significance. A citation from an important document is more important than a citation from a relatively unimportant document. Thus, the importance of a page, and hence the rank assigned to it, should depend not just on the number of citations it has, but on the importance of the citing documents as well. This implies a recursive definition of rank: the rank of a document is a function of the ranks of the documents which cite it. The ranks of documents may be calculated by an iterative procedure on a linked database.

Brief Summary Text (14):
In one aspect of the invention, a computer implemented method is provided for scoring linked database documents. The method comprises the steps of:

Detailed Description Text (3):
A linked database (i.e. any database of documents containing mutual citations, such as the world wide web or other hypermedia archive, a dictionary or thesaurus, and a database of academic articles, patents, or court cases) can be represented as a directed graph of N nodes, where each node corresponds to a web page document and where the directed connections between nodes correspond to links from one document to another. A given node has a set of forward links that connect it to children nodes, and a set of backward links that connect it to parent nodes. FIG. 1 shows a typical relationship between three hypertext documents A, B, and C. As shown in this particular figure, the first links in documents B and C are pointers to document A. In this case we say that B and C are backlinks of A, and that A is a forward link of B and of C. Documents B and C also have other forward links to documents that are not shown.

Detailed Description Text (6):
According to one embodiment of the present method of ranking, the backlinks from different pages are weighted differently and the number of links on each page is normalized. More precisely, the rank of a page A is defined according to the present invention as ##EQU1##

Detailed Description Text (8):
The ranks form a probability distribution over web pages, so that the sum of ranks over all web pages is unity. The rank of a page can be interpreted as the probability that a surfer will be at the page after following a large number of forward links. The constant .alpha. in the formula is interpreted as the probability that the web surfer will jump randomly to any web page instead of following a forward link. The page ranks for all the pages can be calculated using a simple iterative algorithm, and corresponds to the principal eigenvector of the normalized link matrix of the web, as will be discussed in more detail below.

Detailed Description Text (9):
In order to illustrate the present method of ranking, consider the simple web of three documents shown in FIG. 2. For simplicity of illustration, we assume in this example that r=0. Document A has a single backlink to document C, and this is the only forward link of document C, so

Detailed Description Text (13):
Document C has two backlinks. One backlink is to document B, and this is the only forward link of document B. The other backlink is to document A via the other of the two forward links from A. Thus

Detailed Description Text (15):
In this simple illustrative case we can see by inspection that r(A)=0.4, r(B)=0.2, and r(C)=0.4. Although a typical value for .alpha. is .about.0.1, if for simplicity we set .alpha.=0.5 (which corresponds to a 50% chance that a surfer will randomly jump to one of the three pages rather than following a forward link), then the mathematical relationships between the ranks become more complicated. In particular, we then have

Detailed Description Text (21):
The iteration process can be understood as a steady-state probability distribution calculated from a model of a random surfer. This model is mathematically equivalent to the explanation described above, but provides a more direct and concise characterization of the procedure. The model includes (a) an initial N-dimensional probability distribution vector p.sub.0 where each component p.sub.0 [i] gives the initial probability that a random surfer will start at a node i, and (b) an N.times.N transition probability matrix A where each component A[i][j] gives the probability that the surfer will move from node i to node j. The probability distribution of the graph after the surfer follows one link is p.sub.1 =Ap.sub.0, and after two links the probability distribution is p.sub.2 =Ap.sub.1 =A.sup.2

p.sup.0. Assuming this iteration converges, it will converge to a steady-state probability ##EQU2##

Detailed Description Text (22):
which is a dominant eigenvector of A. The iteration circulates the probability through the linked nodes like energy flows through a circuit and accumulates in important places. Because pages with no links occur in significant numbers and bleed off energy, they cause some complication with computing the ranking. This complication is caused by the fact they can add huge amounts to the "random jump" factor. This, in turn, causes loops in the graph to be highly emphasized which is not generally a desirable property of the model. In order to address this problem, these childless pages can simply be removed from the model during the iterative stages, and added back in after the iteration is complete. After the childless pages are added back in, however, the same number of iterations that was required to remove them should be done to make sure they all receive a value. (Note that in order to ensure convergence, the norm of p.sub.i must be made equal to 1 after each iteration.) An alternate method to control the contribution of the childless nodes is to only estimate the steady state by iterating a small number of times.

Detailed Description Text (23):
The rank r[i] of a node i can then be defined as a function of this steady-state probability distribution. For example, the rank can be defined simply by r[i] =p.infin.[i]. This method of calculating rank is mathematically equivalent to the iterative method described first. Those skilled in the art will appreciate that this same method can be characterized in various different ways that are mathematically equivalent. Such characterizations are obviously within the scope of the present invention. Because the rank of various different documents can vary by orders of magnitude, it is convenient to define a logarithmic rank ##EQU3##

Detailed Description Text (24):
which assigns a rank of 0 to the lowest ranked node and increases by 1 for each order of magnitude in importance higher than the lowest ranked node.

Detailed Description Text (25):
"FIG. 3 shows one embodiment of a computer implemented method for calculating an importance rank for N linked nodes of a linked database. At a step 101, an initial N-dimensional vector p.sub.0 is selected. An approximation p.sub.n to a steady-state probability p.sub..infin. in accordance with the equation p.sub.n =A.sup.n p.sub.0 is computed at a step 103. Matrix A can be an N.times.N transition probability matrix having elements A[i][j] representing a probability of moving from node i to node j. At a step 105, a rank r[k] for node k from a k.sup.th component of p.sub.n is determined.".

Detailed Description Text (26):
In one particular embodiment, a finite number of iterations are performed to approximate p.infin.. The initial distribution can be selected to be uniform or non-uniform. A uniform distribution would set each component of p.sub.0 equal to 1/N. A non-uniform distribution, for example, can divide the initial probability among a few nodes which are known a priori to have relatively large importance. This non-uniform distribution decreases the number of iterations required to obtain a close approximation to p.infin. and also is one way to reduce the effect of artificially inflating relevance by adding unrelated terms.

Detailed Description Text (28):
where 1 is an N.times.N matrix consisting of all 1s, .alpha. is the probability that a surfer will jump randomly to any one of the N nodes, and B is a matrix whose elements B[i][j] are given by ##EQU5##

Detailed Description Text (29):
where n.sub.i is the total number of forward links from node i. The (1-.alpha.)

factor acts as a damping factor that limits the extent to which a document's rank can be inherited by children documents. This models the fact that users typically jump to a different place in the web after following a few links. The value of .alpha. is typically around 15%. Including this damping is important when many iterations are used to calculate the rank so that there is no artificial concentration of rank importance within loops of the web. Alternatively, one may set .alpha.=0 and only iterate a few times in the calculation.

Detailed Description Text (30):
Consistent with the present invention, there are several ways that this method can be adapted or altered for various purposes. As already mentioned above, rather than including the random linking probability .alpha. equally among all nodes, it can be divided in various ways among all the sites by changing the 1 matrix to another matrix. For example, it could be distributed so that a random jump takes the surfer to one of a few nodes that have a high importance, and will not take the surfer to any of the other nodes. This can be very effective in preventing deceptively tagged documents from receiving artificially inflated relevance. Alternatively, the random linking probability could be distributed so that random jumps do not happen from high importance nodes, and only happen from other nodes. This distribution would model a surfer who is more likely to make random jumps from unimportant sites and follow forward links from important sites. A modification to avoid drawing unwarranted attention to pages with artificially inflated relevance is to ignore local links between documents and only consider links between separate domains. Because the links from other sites to the document are not directly under the control of a typical web site designer, it is then difficult for the designer to artificially inflate the ranking. A simpler approach is to weight links from pages contained on the same web server less than links from other servers. Also, in addition to servers, internet domains and any general measure of the distance between links could be used to determine such a weighting.

Detailed Description Text (32):
Links can also be weighted by their relative importance within a document. For example, highly visible links that are near the top of a document can be given more weight. Also, links that are in large fonts or emphasized in other ways can be given more weight. In this way, the model better approximates human usage and authors' intentions. In many cases it is appropriate to assign higher value to links coming from pages that have been modified recently since such information is less likely to be obsolete.

Detailed Description Text (38):
Another important application and embodiment of the present invention is directed to enhancing the quality of results from web search engines. In this application of the present invention, a ranking method according to the invention is integrated into a web search engine to produce results far superior to existing methods in quality and performance. A search engine employing a ranking method of the present invention provides automation while producing results comparable to a human maintained categorized system. In this approach, a web crawler explores the web and creates an index of the web content, as well as a directed graph of nodes corresponding to the structure of hyperlinks. The nodes of the graph (i.e. pages of the web) are then ranked according to importance as described above in connection with various exemplary embodiments of the present invention.

CLAIMS:

2. The method of claim 1, wherein the assigning includes:

identifying a weighting factor for each of the linking documents, the weighting factor being dependent on the number of links to the one or more linking documents, and

adjusting the score of each of the one or more linking documents based on the identified weighting factor.

3. The method of claim 1, wherein the assigning includes:

identifying a weighting factor for each of the linking documents, the weighting factor being dependent on an estimation of a probability that a linking document will be accessed, and

adjusting the score of each of the one or more linking documents based on the identified weighting factor.

4. The method of claim 1, wherein the assigning includes:

identifying a weighting factor for each of the linking documents, the weighting factor being dependent on the URL, host, domain, author, institution, or last update time of the one or more linking documents, and

adjusting the score of each of the one or more linking documents based on the identified weighting factor.

5. The method of claim 1, wherein the assigning includes:

identifying a weighting factor for each of the linking documents, the weighting factor being dependent on whether the one or more linking documents are selected documents or roots, and

adjusting the score of each of the one or more linking documents based on the identified weighting factor.

6. The method of claim 1, wherein the assigning includes:

identifying a weighting factor for each of the linking documents, the weighting factor being dependent on the importance, visibility or textual emphasis of the links in the one or more linking documents, and

adjusting the score of each of the one or more linking documents based on the identified weighting factor.

7. The method of claim 1, wherein the assigning includes:

identifying a weighting factor for each of the linking documents, the weighting factor being dependent on a particular user's preferences, the rate at which users access the one or more linking documents, or the importance of the one or more linking documents, and

adjusting the score of each of the one or more linking documents based on the identified weighting factor.

22. The method of claim 1, wherein the assigning a score includes:

associating one or more backlinks with each of the linked documents, each of the backlinks corresponding to one of the linking documents that links to the linked document,

assigning a weight to each of the backlinks, and

determining a score for each of the linked documents based on a number of backlinks for the linked document and the weights assigned to the backlinks.

24. The method of claim 22, wherein the assigning a weight includes:

assigning different weights to at least some of the backlinks associated with at least one of the linked documents.

25. The method of claim 1, wherein the assigning a score includes:

associating one or more backlinks with each of the linked documents, each of the backlinks corresponding to one of the linking documents that links to the linked document,

assigning a weight to each of the backlinks, and

determining a score for each of the linked documents based on a sum of the weights assigned to the backlinks associated with the linked document.

26. The method of claim 25, wherein the weights assigned to each of the backlinks are independent of text of the corresponding linking documents.

Previous Doc Next Doc Go to Doc#

L6: Entry 3 of 11 File: PGPB Sep 30, 2004

DOCUMENT-IDENTIFIER: US 20040193636 A1
TITLE: Method for identifying related pages in a hyperlinked database

Abstract Paragraph:
A method is described for identifying related pages among a plurality of pages in a linked database such as the World Wide Web. An initial page is selected from the plurality of pages. Pages linked to the initial page are represented as a graph in a memory. The pages represented in the graph are scored on content, and a set of pages is selected, the selected set of pages having scores greater than a first predetermined threshold. The selected set of pages is scored on connectivity, and a subset of the set of pages that have scores greater than a second predetermined threshold are selected as related pages.

Summary of Invention Paragraph:
[0006] The vicinity of a Web page is defined by the hyperlinks that connect the page to others. A Web page can point to other pages, and the page can be pointed to by other pages. Close pages are directly linked, farther pages are indirectly linked via intermediate pages. This connectivity can be expressed as a graph where nodes represent the pages, and the directed edges represent the links. The vicinity of all the pages in the result set, up to a certain distance, is called the neighborhood graph.

Summary of Invention Paragraph:
[0011] In U.S. patent application Ser. No. 09/058,577 "Method for Ranking Documents in a Hyperlinked Environment using Connectivity and Selective Content Analysis" filed by Bharat et al. on Apr. 9, 1998, a method is described which performs content analysis on only a small subset of the pages in the neighborhood graph to determine relevance weights, and pages with low relevance weights are pruned from the graph. Then, the pruned graphed is ranked according to a connectivity analysis. This method still requires the result set of a query to form a query topic.

Summary of Invention Paragraph:
[0012] Therefore, there is a need for a method for identifying related pages in a linked database that does not require a query and the fetching of many unrelated pages.

Summary of Invention Paragraph:
[0013] Provided is a method for identifying related pages among a plurality of pages in a linked database such as the World Wide Web. An initial page is selected from the plurality of pages by specifying the URL of the page or clicking on the page using a Web browser in a convenient manner.

Detail Description Paragraph:
[0028] We use the initial page 201 to construct 210 a neighborhood graph (n-graph) 211 in a memory. Nodes 212 in the graph represent the initial selected page 201 as well as other closely linked pages, as described below. The edges 213 denote the hyperlinks between pages. The "size" of the graph is determined by K which can be preset or adjusted dynamically as the graph is constructed. The idea being that the graph needs to represent a meaningful number of page.

Detail Description Paragraph:
[0029] During the construction of the neighborhood graph 211, the direction of links is considered as a way of pruning the graph. In the preferred implementation, with K=2, our method only includes nodes at distance 2 that are reachable by going one link backwards ("B"), pages reachable by going one link forwards ("F"), pages reachable by going one link backwards followed by one link forward ("BF") and those reachable by going one link forwards and one link backwards ("FB"). This eliminates nodes that are reachable only by going forward two links ("FF") or backwards two links ("BB").

Detail Description Paragraph:
[0030] To eliminate some unrelated nodes from the neighborhood graph 211, our method relies on a list 299 of "stop" URLs. Stop URLs are URLs that are so popular that they are frequently referenced from many, many pages, such as, for instance URLs that refer to popular search engines. An example is "www.altavista.com." These "stop" nodes are very general purpose and so are generally not related to the specific topic of the selected page 201, and consequently serve no purpose in the neighborhood graph. Our method checks each URL against the stop list 299 during the neighborhood graph construction, and eliminates the node and all incoming and outgoing edges if a URL is found on the stop list 299.

Detail Description Paragraph:
[0031] In some cases, the neighborhood graph becomes too large. For example, highly popular pages are often pointed to by many thousands of pages and including all such pages in the neighborhood graph is impractical. Similarly, some pages contain thousands of outgoing links, which also cause the graph to become too large. Our method filters the incoming or outgoing edges by choosing only a fixed number M of them. In our preferred implementation, M is 50. In the case that the page was reached by a backwards link L, and the page has more than M outgoing links, our method chooses the M links that surround the link L on the page.

Detail Description Paragraph:
[0033] In some cases, pages will have identical content, or nearly identical contents. This can happen when pages were copied, for example. In such cases, we want to include only one such page in our neighborhood graph, since the presence of multiple copies of a page will tend to artificially increase the importance of any pages that the copies point to. We collapse duplicate pages to a single node in the neighborhood graph. There are several ways that one could identify duplicate pages.

Detail Description Paragraph:
Relevancy Scoring of Nodes in the Neighborhood Graph

Detail Description Paragraph:
[0037] A vector matching operation based on cosine of the angle between vectors is used to produces scores 203 that measure similarity. Please see, Salton et al., "Term-Weighting Approaches in Automatic Text Retrieval," Information Processing and Management, 24(5), 513-23, 1988. A probabilistic model is described by Croft et al. in "Using Probabilistic Models of Document Retrieval without Relevance Feedback," Documentation, 35(4), 285-94, 1979. For a survey of ranking techniques in Information Retrieval see Frakes et al., "Information Retrieval: Data Structures & Algorithms," Chapter 14-`Ranking Algorithms,` Prentice-Hall, N.J., 1992.

Detail Description Paragraph:
[0038] Our topic vector can be determined as the term vector of the initial page 201, or as a vector sum of the term vector of the initial selected page and some function of the term vectors of all the pages presented in the neighborhood graph 211. One such function could simply weight the term vectors of each of the pages equally, while another more complex function would give more weight to the term vectors of pages that are at a smaller distance K from the selected page 201.

Scoring 220 results in a scored graph 215.

Detail Description Paragraph:
Pruning Nodes in the Scored Neighborhood Graph

Detail Description Paragraph:
[0039] After the graph has been scored, the scored graph 215 is "pruned" 230 to produce a pruned graph 216. Here, pruning means removing those nodes and links from the graph that are not "similar." There are a variety of approaches which can be used as the threshold for pruning, including median score, absolute threshold, or a slope-based approach.

Detail Description Paragraph:
[0042] One algorithm which performs this scoring is the Kleinberg algorithm mentioned previously. This algorithm works by iteratively computing two scores for each node in the graph: a hub score (HS) 241 and an authority score 242. The hub score 241 estimates good hub pages, for example, a page such as a directory that points to many other relevant pages. The authority score 242 estimates good authority pages, for example, a page that has relevant information.

Detail Description Paragraph:
[0045] If a single node has dominated the computation as a hub node, that is, exerted "undue influence", then it is sometimes beneficial to remove that node from the neighborhood graph in optional step 250, and repeat the scoring phase 240 on the graph with the node removed. One way of detecting when this undue influence has been exerted is when a single node has a large fraction of the total hub scores of all the nodes (e.g., more than 95% of the total hub scores is attributed to a single node). Another means determines if the node with the highest hub score has more than three times the hub score of the next highest hub score. Other means of determining undue influence are possible.

Detail Description Paragraph:
[0053] Our method differs from Kleinberg's algorithm in the scoring phase in that we detect cases where a node has exerted "undue influence" on the computation of hub scores. In this case, we remove the node from the graph and repeat the scoring computation without this node in the graph. This change tends to produce a more desirable ordering of related pages where highly rated pages are referred to by more than one page. Kleinberg's algorithm does not include any such handling of nodes with undue influence.

CLAIMS:

1. A method for identifying related pages from a plurality of pages in a linked database, comprising the steps of: selecting an initial page from the plurality of pages; representing the initial page and pages linked to the initial page as a graph of nodes and edges in a memory; repeatedly scoring the initial page and the pages linked to the initial page on connectivity of the pages; and selecting a subset of the pages scored on connectivity that have scores greater than a first predetermined threshold as the related pages of the linked database.

11. The method of claim 8 wherein the pages represented in the graph as nodes are linked to the node representing the initial page by a number of edges that is determined dynamically.

14. The method of claim 1 including removing any nodes from the graph that have scores higher than a third predetermined threshold.

16. The method of claim 14 wherein the third predetermined threshold is about three times larger than a next highest scoring node.

L6: Entry 6 of 11 File: USPT Sep 28, 2004

DOCUMENT-IDENTIFIER: US 6799176 B1
TITLE: Method for scoring documents in a linked database

Abstract Text (1):
A method is presented for scoring documents stored in a network. The method includes identifying links from linking documents to linked documents in the network and determining an importance of the identified links. The method further includes weighting the identified links based on the determined importance and scoring the linked documents based on the weighted links.

Brief Summary Text (2):
This invention relates generally to techniques for analyzing linked databases. More particularly, it relates to methods for assigning ranks to nodes in a linked database, such as any database of documents containing citations, the world wide web or any other hypermedia database.

Brief Summary Text (10):
Various aspects of the present invention provide systems and methods for ranking documents in a linked database. One aspect provides an objective ranking based on the relationship between documents. Another aspect of the invention is directed to a technique for ranking documents within a database whose content has a large variation in quality and importance. Another aspect of the present invention is to provide a document ranking method that is scalable and can be applied to extremely large databases such as the world wide web. Additional aspects of the invention will become apparent in view of the following description and associated figures.

Brief Summary Text (11):
One aspect of the present invention is directed to taking advantage of the linked structure of a database to assign a rank to each document in the database, where the document rank is a measure of the importance of a document. Rather than determining relevance only from the intrinsic content of a document, or from the anchor text of backlinks to the document, a method consistent with the invention determines importance from the extrinsic relationships between documents. Intuitively, a document should be important (regardless of its content) if it is highly cited by other documents. Not all citations, however, are necessarily of equal significance. A citation from an important document is more important than a citation from a relatively unimportant document. Thus, the importance of a page, and hence the rank assigned to it, should depend not just on the number of citations it has, but on the importance of the citing documents as well. This implies a recursive definition of rank: the rank of a document is a function of the ranks of the documents which cite it. The ranks of documents may be calculated by an iterative procedure on a linked database.

Brief Summary Text (13):
In one aspect of the invention, a computer implemented method is provided for scoring linked documents. The method includes identifying links from linking documents to linked documents in the network and determining an importance of the identified links. The method further includes weighting the identified links based on the determined importance and scoring the linked documents based on the weighted links.

Brief Summary Text (15):
In accordance with yet another implementation consistent with the present invention, a method scoring documents stored in a network includes traversing the network to identify links between the documents; identifying a location at which each of the documents is stored; weighting the links between documents based on the identified locations; and scoring the documents based on the weighted links.

Brief Summary Text (16):
In accordance with a further implementation consistent with the present invention, a method for scoring documents stored in a network includes identifying links from linking documents to linked documents in the network; determining an importance of the identified links; weighting the identified links based on the determined importance; and scoring the linked documents based on the weighted links.

Brief Summary Text (20):
In yet another implementation consistent with the present invention, a method of organizing linked documents includes: (a) identifying a first linked document; (b) identifying links between linking documents and the first linked document; (c) assigning a weight to each of the identified links; (d) determining a score for the first linked document based on (i) the identified links between the linking documents and the first linked document, and (ii) the weights assigned to each of the identified links; (e) repeating steps (a)-(d) for a second linked document; and (f) organizing the first and second linked documents based on the determined scores.

Detailed Description Text (3):
A linked database (i.e. any database of documents containing mutual citations, such as the world wide web or other hypermedia archive, a dictionary or thesaurus, and a database of academic articles, patents, or court cases) can be represented as a directed graph of N nodes, where each node corresponds to a web page document and where the directed connections between nodes correspond to links from one document to another. A given node has a set of forward links that connect it to children nodes, and a set of backward links that connect it to parent nodes. FIG. 1 shows a typical relationship between three hypertext documents A, B, and C. As shown in this particular figure, the first links in documents B and C are pointers to document A. In this case we say that B and C are backlinks of A, and that A is a forward link of B and of C. Documents B and C also have other forward links to documents that are not shown.

Detailed Description Text (5):
According to one embodiment of the present method of ranking, the backlinks from different pages are weighted differently and the number of links on each page is normalized. More precisely, the rank of a page A is defined according to the present invention as ##EQU1##

Detailed Description Text (7):
The ranks form a probability distribution over web pages, so that the sum of ranks over all web pages is unity. The rank of a page can be interpreted as the probability that a surfer will be at the page after following a large number of forward links. The constant .alpha. in the formula is interpreted as the probability that the web surfer will jump randomly to any web page instead of following a forward link. The page ranks for all the pages can be calculated using a simple iterative algorithm, and corresponds to the principal eigenvector of the normalized link matrix of the web, as will be discussed in more detail below.

Detailed Description Text (8):
In order to illustrate the present method of ranking, consider the simple web of three documents shown in FIG. 2. For simplicity of illustration, we assume in this example that r=0. Document A has a single backlink to document C, and this is the

only forward link of document C, so

Detailed Description Text (10):
Document C has two backlinks. One backlink is to document B, and this is the only forward link of document B. The other backlink is to document A via the other of the two forward links from A. Thus

Detailed Description Text (11):
In this simple illustrative case we can see by inspection that r(A)=0.4, r(B)=0.2, and r(C)=0.4. Although a typical value for .alpha. is .about.0.1, if for simplicity we set .alpha.=0.5 (which corresponds to a 50% chance that a surfer will randomly jump to one of the three pages rather than following a forward link), then the mathematical relationships between the ranks become more complicated. In particular, we then have

Detailed Description Text (15):
The iteration process can be understood as a steady-state probability distribution calculated from a model of a random surfer. This model is mathematically equivalent to the explanation described above, but provides a more direct and concise characterization of the procedure. The model includes (a) an initial N-dimensional probability distribution vector p.sub.0 where each component p.sub.0 [i] gives the initial probability that a random surfer will start at a node i, and (b) an N.times.N transition probability matrix A where each component A[i][j] gives the probability that the surfer will move from node i to node j. The probability distribution of the graph after the surfer follows one link is p.sub.1 =Ap.sub.0, and after two links the probability distribution is p.sub.2 =Ap.sub.1 =A.sup.2 p.sub.0. Assuming this iteration converges, it will converge to a steady-state probability ##EQU2##

Detailed Description Text (16):
which is a dominant eigenvector of A. The iteration circulates the probability through the linked nodes like energy flows through a circuit and accumulates in important places. Because pages with no links occur in significant numbers and bleed off energy, they cause some complication with computing the ranking. This complication is caused by the fact they can add huge amounts to the "random jump" factor. This, in turn, causes loops in the graph to be highly emphasized which is not generally a desirable property of the model. In order to address this problem, these childless pages can simply be removed from the model during the iterative stages, and added back in after the iteration is complete. After the childless pages are added back in, however, the same number of iterations that was required to remove them should be done to make sure they all receive a value. (Note that in order to ensure convergence, the norm of p.sub.i must be made equal to 1 after each iteration.) An alternate method to control the contribution of the childless nodes is to only estimate the steady state by iterating a small number of times.

Detailed Description Text (17):
The rank r[i] of a node i can then be defined as a function of this steady-state probability distribution. For example, the rank can be defined simply by r[i] =p.sub..infin. [i]. This method of calculating rank is mathematically equivalent to the iterative method described first. Those skilled in the art will appreciate that this same method can be characterized in various different ways that are mathematically equivalent. Such characterizations are obviously within the scope of the present invention. Because the rank of various different documents can vary by orders of magnitude, it is convenient to define a logarithmic rank ##EQU3##

Detailed Description Text (18):
which assigns a rank of 0 to the lowest ranked node and increases by 1 for each order of magnitude in importance higher than the lowest ranked node.

Detailed Description Text (19):

FIG. 3 shows one embodiment of a computer implemented method for calculating an importance rank for N linked nodes of a linked database. At a step 101, an initial N-dimensional vector p.sub.0 is selected. An approximation p.sub.n to a steady-state probability p.sub..infin. in accordance with the equation p.sub.n =A.sup.n p.sub.0 is computed at a step 103. Matrix A can be an N.times.N transition probability matrix having elements A[i][j] representing a probability of moving from node i to node j. At a step 105, a rank r[k] for node k from a k.sup.th component of p.sub.n is determined.

Detailed Description Text (20):
In one particular embodiment, a finite number of iterations are performed to approximate p.sub..infin.. The initial distribution can be selected to be uniform or non-uniform. A uniform distribution would set each component of p.sub.0 equal to 1/N. A non-uniform distribution, for example, can divide the initial probability among a few nodes which are known a priori to have relatively large importance. This non-uniform distribution decreases the number of iterations required to obtain a close approximation to p.sub..infin. and also is one way to reduce the effect of artificially inflating relevance by adding unrelated terms.

Detailed Description Text (22):
where 11 is an N.times.N matrix consisting of all 1s, .alpha. is the probability that a surfer will jump randomly to any one of the N nodes, and B is a matrix whose elements B[i][J] are given by ##EQU4##

Detailed Description Text (23):
where n.sub.i is the total number of forward links from node i. The (1-.alpha.) factor acts as a damping factor that limits the extent to which a document's rank can be inherited by children documents. This models the fact that users typically jump to a different place in the web after following a few links. The value of .alpha. is typically around 15%. Including this damping is important when many iterations are used to calculate the rank so that there is no artificial concentration of rank importance within loops of the web. Alternatively, one may set .alpha.=0 and only iterate a few times in the calculation.

Detailed Description Text (24):
Consistent with the present invention, there are several ways that this method can be adapted or altered for various purposes. As already mentioned above, rather than including the random linking probability .alpha. equally among all nodes, it can be divided in various ways among all the sites by changing the 11 matrix to another matrix. For example, it could be distributed so that a random jump takes the surfer to one of a few nodes that have a high importance, and will not take the surfer to any of the other nodes. This can be very effective in preventing deceptively tagged documents from receiving artificially inflated relevance. Alternatively, the random linking probability could be distributed so that random jumps do not happen from high importance nodes, and only happen from other nodes. This distribution would model a surfer who is more likely to make random jumps from unimportant sites and follow forward links from important sites. A modification to avoid drawing unwarranted attention to pages with artificially inflated relevance is to ignore local links between documents and only consider links between separate domains. Because the links from other sites to the document are not directly under the control of a typical web site designer, it is then difficult for the designer to artificially inflate the ranking. A simpler approach is to weight links from pages contained on the same web server less than links from other servers. Also, in addition to servers, internet domains and any general measure of the distance between links could be used to determine such a weighting.

Detailed Description Text (26):
Links can also be weighted by their relative importance within a document. For example, highly visible links that are near the top of a document can be given more weight. Also, links that are in large fonts or emphasized in other ways can be

given more weight. In this way, the model better approximates human usage and authors' intentions. In many cases it is appropriate to assign higher value to links coming from pages that have been modified recently since such information is less likely to be obsolete.

Detailed Description Text (31):
Another important application and embodiment of the present invention is directed to enhancing the quality of results from web search engines. In this application of the present invention, a ranking method according to the invention is integrated into a web search engine to produce results far superior to existing methods in quality and performance. A search engine employing a ranking method of the present invention provides automation while producing results comparable to a human maintained categorized system. In this approach, a web crawler explores the web and creates an index of the web content, as well as a directed graph of nodes corresponding to the structure of hyperlinks. The nodes of the graph (i.e., pages of the web) are then ranked according to importance as described above in connection with various exemplary embodiments of the present invention.

Other Reference Publication (2):
Copy of claims of U.S. Serial No. 09/895,174, filed on July 2, 2001; Lawrence Page; Method for Node Ranking in a Linked Database; 8 pages.

Previous Doc Next Doc Go to Doc#

L6: Entry 10 of 11 File: USPT Oct 24, 2000

DOCUMENT-IDENTIFIER: US 6138113 A
TITLE: Method for identifying near duplicate pages in a hyperlinked database

Abstract Text (1):
A method is described for identifying pages that are near duplicates in a linked database. In the linked database, pages can have incoming links and outgoing links. Two pages are selected, a first page and a second page. For each selected page, the number of outgoing links is determined. The two pages are marked as near duplicates based on the number of common outgoing links for the two pages.

Brief Summary Text (8):
The vicinity of a Web page is defined by the hyperlinks that connect the page to others. A Web page can point to other pages, and the page can be pointed to by other pages. Close pages are directly linked, farther pages are indirectly linked. This connectivity can be expressed as a graph where nodes represent the pages, and the directed edges represent the links. The vicinity of all the pages in the result set is called the neighborhood graph.

Brief Summary Text (11):
In U.S. patent application Ser. No. 09/058,577 "Method for Ranking Documents in a Hyperlinked Environment using Connectivity and Selective Content Analysis" filed by Bharat et al. on Apr. 9, 1998, a method is described which performs content analysis only a small subset of the pages in the neighborhood graph to determine relevance weights, and pages with low relevance weights are pruned from the graph. Then, the pruned graphed is ranked according to a connectivity analysis. This method still requires the result set of a query to form a query topic.

Brief Summary Text (14):
Provided is a method for identifying near duplicate pages among a plurality of pages in a linked database such as the World Wide Web. A first and second page are selected for a near duplicate determination. For each page, the number of outgoing links is counted. Pages are marked as near duplicates based on the number of common outgoing links between the two pages.

Detailed Description Text (16):
We use the initial page 201 to construct 210 a neighborhood graph (ngraph) 211 in a memory. Nodes 212 in the graph represent the initial selected page 201 as well as other closely linked pages, as described below. The edges 213 denote the hyperlinks between pages. The "size" of the graph is determined by K which can be preset or adjusted dynamically as the graph is constructed. The idea being that the graph needs to represent a meaningful number of page.

Detailed Description Text (17):
During the construction of the neighborhood graph 211, the direction of links is considered as a way of pruning the graph. In the preferred implementation, with K=2, our method only includes nodes at distance 2 that are reachable by going one link backwards ("B"), pages reachable by going one link forwards ("F"), pages reachable by going one link backwards followed by one link forward ("BF") and those reachable by going one link forwards and one link backwards ("FB"). This eliminates nodes that are reachable only by going forward two links ("FF") or backwards two

links ("BB").

Detailed Description Text (18):
To eliminate some unrelated nodes from the neighborhood graph 211, our method relies on a list 299 of "stop" URLs, which are URLs that are so popular that they are highly referenced from many, many pages, such as popular search engines. An example is "www.altavista.digital.com." These "stop" nodes are very general purpose and so are generally not related to the specific topic of the selected page 201, and so serve no purpose in the neighborhood graph. Our method checks each URL against the stop list 299 during the neighborhood graph construction, and eliminates the node and all incoming and outgoing edges if a URL is found on the stop list 299.

Detailed Description Text (19):
In some cases, the neighborhood graph becomes too large. For example, highly popular pages are often pointed to by many thousands of pages and including all such pages in the neighborhood graph is impractical. Similarly, some pages contain thousands of outgoing links, which also cause the graph to become too large. Our method filters the incoming or outgoing edges by choosing only a fixed number M of them. In our preferred implementation, M is 50. In the case that the page was reached by a backwards link L, and the page has more than M outgoing links, our method chooses the M links that surround the link L on the page.

Detailed Description Text (21):
In some cases, two pages will have identical contents, or nearly identical contents. This can happen when the page was copied, for example. In such cases, we want to include only one such page in our neighborhood graph, since the presence of multiple copies of a page will tend to artificially increase the importance of any pages that they point to. We collapse duplicate pages to a single node in the neighborhood graph. There are several ways that one could identify duplicate pages.

Detailed Description Text (23):
Relevancy Scoring of Nodes in the Neighborhood Graph

Detailed Description Text (25):
Scoring can be done using well known retrieval techniques. For example, in the Salton & Buckley model, the content of the represented pages 211 and the topic 202 can be regarded as vectors in an n-dimensional vector space, where n corresponds to the number of unique terms in the data set. A vector matching operation based on cosine of the angle between vectors is used to produces scores 203 that measure similarity. Please see, Salton et al., "Term-Weighting Approaches in Automatic Text Retrieval,", Information Processing and Management, 24(5), 513-23, 1988. A probabilistic model is described by Croft et al. in "Using Probabilistic Models of Document Retrieval without Relevance Feedback," Documentation, 35(4), 285-94, 1979. For a survey of ranking techniques in Information Retrieval see Frakes et al., "Information Retrieval: Data Structures & Algorithms," Chapter 14--`Ranking Algorithms,` Prentice-Hall, N.J., 1992.

Detailed Description Text (26):
Our topic vector can be determined as the term vector of the initial page 201, or as a vector sum of the term vector of the initial selected page and some function of the term vectors of all the pages presented in the neighborhood graph 211. One such function could simply weight the term vectors of each of the pages equally, while another more complex function would give more weight to the term vectors of pages that are at a smaller distance K from the selected page 201. Scoring 220 results in a scored graph 215.

Detailed Description Text (27):
Pruning Nodes in the Scored Neighborhood Graph

Detailed Description Text (28):
After the graph has been scored, the scored graph 215 is "pruned" 230 to produce a pruned graph 216. Here, pruning means removing those nodes and links from the graph that are not "similar." There are a variety of approaches which can be used as the threshold for pruning, including median score, absolute threshold, or a slope-based approach.

Detailed Description Text (31):
One algorithm which performs this scoring is the Kleinberg algorithm mentioned previously. This algorithm works by iteratively computing two scores for each node in the graph: a hub score (HS) 241 and an authority score 242. The hub score 241 estimates good hub pages, for example, a page such as a directory that points to many other relevant pages. The authority score 242 estimates good authority pages, for example, a page that has relevant information.

CLAIMS:

1. A method for identifying pages that are near duplicates in a linked database, the pages in the database having incoming links and outgoing links, comprising the steps of:

selecting a first page and a second page;

determining the outgoing links for the first page and the second page;

determining the number of outgoing links that are common for the first page and the second page;

marking the first page and the second page as near duplicate pages based on the number of common outgoing links.

Previous Doc Next Doc Go to Doc#

Hit List

First Hit | Clear | Generate Collection | Print | Fwd Refs | Bkwd Refs | Generate OACS

Search Results - Record(s) 1 through 11 of 11 returned.

---

1. Document ID: US 20050240621 A1

Using default format because multiple data bases are involved.

L6: Entry 1 of 11 File: PGPB Oct 27, 2005

PGPUB-DOCUMENT-NUMBER: 20050240621
PGPUB-FILING-TYPE: new
DOCUMENT-IDENTIFIER: US 20050240621 A1

TITLE: Method and system for managing partitioned data resources

PUBLICATION-DATE: October 27, 2005

INVENTOR-INFORMATION:

NAME	CITY	STATE	COUNTRY
Robertson, James A.	The Woodlands	TX	US
Greene, William S.	Fairview	TX	US

US-CL-CURRENT: 707/102

Full | Title | Citation | Front | Review | Classification | Date | Reference | Sequences | Attachments | Claims | KWIC | Draw De

---

2. Document ID: US 20050033742 A1

L6: Entry 2 of 11 File: PGPB Feb 10, 2005

PGPUB-DOCUMENT-NUMBER: 20050033742
PGPUB-FILING-TYPE: new
DOCUMENT-IDENTIFIER: US 20050033742 A1

TITLE: Methods for ranking nodes in large directed graphs

PUBLICATION-DATE: February 10, 2005

INVENTOR-INFORMATION:

NAME	CITY	STATE	COUNTRY
Kamvar, Sepandar D.	Palo Alto	CA	US
Haveliwala, Taher H.	Mountain View	CA	US
Jeh, Glen	San Francisco	CA	US
Golub, Gene	Stanford	CA	US

US-CL-CURRENT: 707/7

Full | Title | Citation | Front | Review | Classification | Date | Reference | Sequences | Attachments | Claims | KWIC | Draw De

---

☐ 3. Document ID: US 20040193636 A1

L6: Entry 3 of 11 File: PGPB Sep 30, 2004

PGPUB-DOCUMENT-NUMBER: 20040193636
PGPUB-FILING-TYPE: new
DOCUMENT-IDENTIFIER: US 20040193636 A1

TITLE: Method for identifying related pages in a hyperlinked database

PUBLICATION-DATE: September 30, 2004

INVENTOR-INFORMATION:

NAME	CITY	STATE	COUNTRY
Black, Jeffrey Dean	Laural	MD	US
Henzinger, Monika R.	Menlo Park	CA	US
Broder, Andrei Z.	Menlo Park	CA	US

US-CL-CURRENT: 707/102

Full | Title | Citation | Front | Review | Classification | Date | Reference | Sequences | Attachments | Claims | KWIC | Draw De

---

☐ 4. Document ID: US 20020165727 A1

L6: Entry 4 of 11 File: PGPB Nov 7, 2002

PGPUB-DOCUMENT-NUMBER: 20020165727
PGPUB-FILING-TYPE: new
DOCUMENT-IDENTIFIER: US 20020165727 A1

TITLE: Method and system for managing partitioned data resources

PUBLICATION-DATE: November 7, 2002

INVENTOR-INFORMATION:

NAME	CITY	STATE	COUNTRY
Greene, William S.	Fairview	TX	US
Robertson, James A.	The Woodlands	TX	US

US-CL-CURRENT: 705/1; 707/1

Full | Title | Citation | Front | Review | Classification | Date | Reference | Sequences | Attachments | Claims | KWIC | Draw De

---

☐ 5. Document ID: US 6922685 B2

L6: Entry 5 of 11 File: USPT Jul 26, 2005

US-PAT-NO: 6922685
DOCUMENT-IDENTIFIER: US 6922685 B2

TITLE: Method and system for managing partitioned data resources

DATE-ISSUED: July 26, 2005

INVENTOR-INFORMATION:

NAME	CITY	STATE	ZIP CODE	COUNTRY
Greene; William S.	Fairview	TX		
Robertson; James A.	The Woodlands	TX		

ASSIGNEE-INFORMATION:

NAME	CITY	STATE	ZIP CODE	COUNTRY	TYPE CODE
MCI, Inc.	Ashburn	VA			02

APPL-NO: 09/863456 [PALM]
DATE FILED: May 22, 2001

PARENT-CASE:
CROSS REFERENCES TO RELATED APPLICATIONS The present application claims the benefit of priority, of U.S. patent application 60/206,564 entitled "METHOD AND SYSTEM FOR MANAGING PARTITIONED DATA RESOURCES," and filed on May 22, 2000. The above identified application is incorporated by reference in its entirety.

INT-CL: [07] G06 F 17/30

US-CL-ISSUED: 707/1; 719/316
US-CL-CURRENT: 707/1; 719/316

FIELD-OF-SEARCH: 707/1-10, 707/100-104.1, 707/103R, 707/103Y, 707/103Z, 719/310-320

PRIOR-ART-DISCLOSED:

U.S. PATENT DOCUMENTS

PAT-NO	ISSUE-DATE	PATENTEE-NAME	US-CL
5019961	May 1991	Addesso et al.	700/87
5515531	May 1996	Fujiwara et al.	
5560006	September 1996	Layden et al.	
5680530	October 1997	Selfridge et al.	
5721913	February 1998	Ackroff et al.	
5774661	June 1998	Chatterjee et al.	
5779297	July 1998	Goodridge et al.	
5848393	December 1998	Goodridge et al.	
5940804	August 1999	Turley et al.	
5960194	September 1999	Choy et al.	
5999910	December 1999	Goodridge et al.	
6006193	December 1999	Gibson et al.	
6092061	July 2000	Choy	
6141773	October 2000	St. Pierre et al.	714/20

6167449	December 2000	Arnold et al.	
6223182	April 2001	Agarwal et al.	
6385626	May 2002	Tamer et al.	707/203
6536935	March 2003	Parunak et al.	
6564216	May 2003	Waters	707/10
6574633	June 2003	Jamalabad et al.	
6611822	August 2003	Beams et al.	
6611840	August 2003	Baer et al.	
6640249	October 2003	Bowman-Amuah	
6772131	August 2004	Francis et al.	
2001/0043600	November 2001	Chatterjee et al.	
2002/0091703	July 2002	Bayles	
2002/0091827	July 2002	King et al.	
2002/0169890	November 2002	Beaumont et al.	
2002/0188720	December 2002	Terrell et al.	
2003/0140119	July 2003	Archarya et al.	
2004/0003058	January 2004	Trossen	

OTHER PUBLICATIONS

Orfali et al., "The Essential Distributed Objects Survival Guide", John Wiley & Sons, Inc., Chapter 19, pp. 325-338, published 1996.
U.S. Appl. No. 60/206,564, Greene, et al., filed May 22, 2000.
"Snap Using the Snap Communication Component", Template Software, 1998, Whole Manual.
Goldbarb et al., "The XML Handbook", 1998, pp. 3-18, 101-118.
"Using the Web Component", Template Software, 1997, Chapters 1-3.
"Workflow Template Training Course", Version 8.0, 1997, Section A.
"Developing a WFt Workflow System," Template Software, 1998, Whole Manual.
"Using the WFT Development Environment", Template Software, 1998, Whole Manual.
"M2 Presswire, Template & Winstar Deploy First Customer Centric Servies Ordering & Provisioning System", Business Wire, Oct. 21, 1998.
"Template Software and Winstar Telecommunications Announce Deployment of Industry's First Customer Centric Ordering & Provisioning System", Business Wire, Oct. 5, 1998.

ART-UNIT: 2126

PRIMARY-EXAMINER: Courtenay, III; St. John

ABSTRACT:

In accordance with an exemplary embodiment of the present invention, association forming entities are: a) maintained as objects in a like manner to the data objects being associated; and b) are themselves partitioned objects comprising two or more association fragments, each association fragment being mostly concerned with the interfaces to a particular data object participating in the association. In accordance with an exemplary embodiment of the present invention, each association fragment affiliated with a particular data object is stored in a location that enhances the ease of interaction between the association fragment and the data object. For example, where a first data object and second data object are maintained in data stores at some distance from one another, physically or logically, then a first association fragment will be located with or near to the

first data object, and a second association fragment will be located with or near the second data object, at least within the same partition. This arrangement may be preferable because the volume of interaction between a data object and its respective association fragment may far outweigh the interaction needed between the two association fragments. This arrangement may also be preferable as the volume of interaction between a client application and both the data object and respective association fragment may exceed the interaction needed between the two association fragments. Some interactions will employ only one of the association fragments with the net result being a reduction in communications requirements and an improvement in performance. The present invention further provides for defining logical domains which are arbitrary and entirely orthogonal to partitions.

64 Claims, 50 Drawing figures

Full | Title | Citation | Front | Review | Classification | Date | Reference | Claims | KWIC | Draw De

---

6. Document ID: US 6799176 B1

L6: Entry 6 of 11 File: USPT Sep 28, 2004

US-PAT-NO: 6799176
DOCUMENT-IDENTIFIER: US 6799176 B1

TITLE: Method for scoring documents in a linked database

DATE-ISSUED: September 28, 2004

INVENTOR-INFORMATION:

NAME	CITY	STATE	ZIP CODE	COUNTRY
Page; Lawrence	Stanford	CA		

ASSIGNEE-INFORMATION:

NAME	CITY	STATE	ZIP CODE	COUNTRY	TYPE CODE
The Board of Trustees of the Leland Stanford Junior University	Palo Alto	CA			02

APPL-NO: 09/899068 [PALM]
DATE FILED: July 6, 2001

PARENT-CASE:
CROSS-REFERENCES TO RELATED APPLICATIONS This application is a continuation of application Ser. No. 09/004,827, filed Jan. 9, 1998, now U.S. Pat. No. 6,285,999. This application claims priority from U.S. provisional patent application No. 60/035,205 filed Jan. 10, 1997, which is incorporated herein by reference.

INT-CL: [07] G06 F 17/30

US-CL-ISSUED: 707/5; 715/501.1
US-CL-CURRENT: 707/5; 715/501.1

FIELD-OF-SEARCH: 707/5, 707/7, 707/10, 707/1-3, 715/501.1, 702/179, 702/181

PRIOR-ART-DISCLOSED:

U.S. PATENT DOCUMENTS

PAT-NO	ISSUE-DATE	PATENTEE-NAME	US-CL
4953106	August 1990	Gansner et al.	
5450535	September 1995	North	
5594897	January 1997	Goffman	707/102
5748954	May 1998	Mauldin	
5752241	May 1998	Cohen	
5754939	May 1998	Herz et al.	455/3.04
5832494	November 1998	Egger et al.	
5848407	December 1998	Ishikawa et al.	
5915249	June 1999	Spencer	
5920854	July 1999	Kirsch et al.	
5920859	July 1999	Li	707/5
6014678	January 2000	Inoue et al.	
6112202	August 2000	Kleinberg	707/5
6163778	December 2000	Fogg et al.	707/10
6269368	July 2001	Diamond	707/6
6285999	September 2001	Page	707/5
6389436	May 2002	Chakrabarti et al.	707/513
2001/0002466	May 2001	Krasle	704/270.1

OTHER PUBLICATIONS

Recker et al "Predicting document access in large multimedia repositories", ACM Transactions on Computer-Human Interaction, vol. 3, No. 4, Dec. 1996, pp. 352-375.*
Copy of claims of U.S. Serial No. 09/895,174, filed on July 2, 2001; Lawrence Page; Method for Node Ranking in a Linked Database; 8 pages.
Yuwono et al., "Search and Ranking Algorithms for Locating Resources on the World Wide Web", IEEE 1996, pp. 164-171.
L. Katz, "A new status index derived from sociometric analysis", 1953, Psychometricka, vol. 18, pp. 39-43.
C.H. Hubbell, "An input-output approach to clique identification sociometry", 1965, pp. 377-399.
Mizruchi et al., "Techniques for disaggregation centrality scores in social networks", 1996, Sociological Methodology, pp. 26-48.
E. Garfield, "Citation analysis as a tool in journal evaluation", 1972, Science, vol. 178, pp. 471-479.
Pinski et al., "Citation influence for journal aggregates of scientific publications: Theory, with application to the literature of physics", 1976, Inf. Proc. and Management, vol. 12, pp. 297-312.
N. Geller, "On the citation influence methodology of Pinski and Narin", 1978, Inf. Proc. and Management, vol. 14, pp. 93-95.
P. Doreian, "Measuring the relative standing of disciplinary journals", 1988, Inf. Proc. and Management, vol. 24, pp. 45-56.
P. Doreian, "A measure of standing for citation networks within a wider environment", 1994, Inf. Proc. and Management, vol. 30, pp. 21-31.
Botafogo et al., "Structural analysis of hypertext: Identifying hierarchies and useful metrics", 1992, ACM Trans. Inc. Systems, vol. 10, pp. 142-180.
Mark E. Frisse, "Searching for information in a hypertext medical handbook", 1988, Communications of the ACM, vol. 31, No. 7, pp. 880-886.
Massimo Marchiori, "The quest for correct information on the Web: Hyper search engines", 1997 Computer Networks and ISDN Systems, vol. 29, No. 8-13, pp. 1225-

1235. Printed May 28, 2002 from the Internet.
Oliver A. McBryan, "GENVL and WWWW: Tools for taming the web", 1994, Proceedings of the first International World Wide Web Conference, pp. 1-13.
Carriere et al., "Web Query: Searching and visualizing the web through connectivity", 1997, Proc. 6.sup.th International World Wide Web Conference, pp. 1-14. Printed Aug. 15, 2000 from the Internet.
Arocena et al., "Applications of a web query language", 1997, Computer Networks and ISDN Systems, vol. 29, No. 8-13, pp. 1305-1316. Printed May 28, 2002 From the Internet.
Jon M. Kleinberg, "Authoritative sources in a hyperlinked environment", 1998, Proc. of the 9.sup.th Annual ACM-SIAM Symposium on Discrete Algorithms, pp. 668-677.
Henzinger et al., "Measuring index quality using random walks on the web", 1999, Proc. of the 8.sup.th International World Wide Web Conference, pp. 213-225.
Craig Boyle, "To link or not to link: An empirical comparison of Hypertext linking strategies", ACM, 1992, pp. 221-231.
S. Jeromy Carriere et al., "Web Query: Search and Visualizing the Web through Connectivity", Computer Networks and ISDN Systems 29 (1997), pp. 1257-1267.
Wang et al., "Prefetching in World Wide Web", IEEE 1996, pp. 28-32.
Ramer et al., "Similarity, Probability and Database Organisation: Extended Abstract", IEEE 1996, pp. 272-276.
J. Kleinberg, "Authoritative Sources in a Hyperlinked Environment," IBM Research Report, RJ 10076 (91892), May 29, 1997, 37 pages.

ART-UNIT: 2171

PRIMARY-EXAMINER: Le; Uyen

ATTY-AGENT-FIRM: Harrity & Snyder, LLP

ABSTRACT:

A method is presented for scoring documents stored in a network. The method includes identifying links from linking documents to linked documents in the network and determining an importance of the identified links. The method further includes weighting the identified links based on the determined importance and scoring the linked documents based on the weighted links.

1 Claims, 3 Drawing figures

☐ 7. Document ID: US 6665837 B1

L6: Entry 7 of 11 File: USPT Dec 16, 2003

US-PAT-NO: 6665837
DOCUMENT-IDENTIFIER: US 6665837 B1

TITLE: Method for identifying related pages in a hyperlinked database

DATE-ISSUED: December 16, 2003

INVENTOR-INFORMATION:

NAME	CITY	STATE	ZIP CODE	COUNTRY

Dean; Jeffrey	Menlo Park	CA
Henzinger; Monika R.	Menlo Park	CA
Broder; Andrei Z.	Menlo Park	CA

ASSIGNEE-INFORMATION:

NAME	CITY	STATE	ZIP CODE	COUNTRY	TYPE CODE
Overture Services, Inc.	Pasadena	CA			02

APPL-NO: 09/131473 [PALM]
DATE FILED: August 10, 1998

INT-CL: [07] G06 F 15/00

US-CL-ISSUED: 715/501.1; 715/513, 707/2
US-CL-CURRENT: 715/501.1; 707/2, 715/513

FIELD-OF-SEARCH: 707/2, 707/4, 707/5, 707/102, 707/513, 707/501, 715/501.1, 715/513

PRIOR-ART-DISCLOSED:

U.S. PATENT DOCUMENTS

PAT-NO	ISSUE-DATE	PATENTEE-NAME	US-CL
5418948	May 1995	Turle	707/4
5594897	January 1997	Goffman	707/102
5724567	March 1998	Rose et al.	707/2
5855015	December 1998	Shoham	707/5
5895470	April 1999	Pirolli et al.	707/102
5905863	May 1999	Knowles et al.	707/501
5933823	August 1999	Cullen et al.	707/6
5991713	November 1999	Unger et al.	707/513
6073135	June 2000	Broder et al.	707/100
6112202	August 2000	Kleinberg	707/5
6112203	August 2000	Bharat et al.	707/5
6115718	September 2000	Huberman et al.	707/102
6138113	October 2000	Dean et al.	707/2
6334145	December 2001	Adams et al.	709/217

OTHER PUBLICATIONS

Guinan et al., Infromation Retrieval from Hypertext Using Dynamically Planned Guided Tours, ACM 1992, pp. 122-130.*
Salton et al., Selective Text Utilization and Text Traversal, ACM 1993, pp. 131-144.*
Chekuri et al, Web Search Using Automatic Classification, Google, Dec. 1996, pp. 1-11.*
Kleinberg, Authoritative Sources in a Hyperlinked Environment, Google, May 1997, pp. 668-677.

ART-UNIT: 2178

PRIMARY-EXAMINER: Hong; Stephen S.

ASSISTANT-EXAMINER: Huynh; Cong-Lac

ATTY-AGENT-FIRM: Brown Raysman Millstein Felder & Steiner LLP

ABSTRACT:

A method is described for identifying related pages among a plurality of pages in a linked database such as the World Wide Web. An initial page is selected from the plurality of pages. Pages linked to the initial page are represented as a graph in a memory. The pages represented in the graph are scored on content, and a set of pages is selected, the selected set of pages having scores greater than a first predetermined threshold. The selected set of pages is scored on connectivity, and a subset of the set of pages that have scores greater than a second predetermined threshold are selected as related pages.

54 Claims, 2 Drawing figures

Full | Title | Citation | Front | Review | Classification | Date | Reference | Claims | KWIC | Draw. D

---

8. Document ID: US 6578078 B1

L6: Entry 8 of 11 File: USPT Jun 10, 2003

US-PAT-NO: 6578078
DOCUMENT-IDENTIFIER: US 6578078 B1

TITLE: Method for preserving referential integrity within web sites

DATE-ISSUED: June 10, 2003

INVENTOR-INFORMATION:

NAME	CITY	STATE	ZIP CODE	COUNTRY
Smith; Michael D.	Kirkland	WA		
Hennings; Eric	Seattle	WA		
McKee; Christine Walpole	Seattle	WA		

ASSIGNEE-INFORMATION:

NAME	CITY	STATE	ZIP CODE	COUNTRY	TYPE CODE
Microsoft Corporation	Redmond	WA			02

APPL-NO: 09/285530 [PALM]
DATE FILED: April 2, 1999

INT-CL: [07] G06 F 17/30, G06 F 15/173

US-CL-ISSUED: 709/224; 707/513, 707/10
US-CL-CURRENT: 709/224; 707/10, 715/513

FIELD-OF-SEARCH: 709/201, 709/202, 709/203, 709/217, 709/218, 709/219, 709/224, 709/225, 709/226, 709/229, 707/500.1, 707/501.1, 707/513, 707/10

PRIOR-ART-DISCLOSED:

U.S. PATENT DOCUMENTS

PAT-NO	ISSUE-DATE	PATENTEE-NAME	US-CL
5761683	June 1998	Logan et al.	707/513
6240455	May 2001	Kamasaka et al.	709/200
6253204	June 2001	Glass et al.	707/10
6321242	November 2001	Fogg et al.	707/10

ART-UNIT: 2153

PRIMARY-EXAMINER: Dinh; Dung C.

ATTY-AGENT-FIRM: Anderson; Ronald M.

ABSTRACT:

The integrity of uniform resource locator (URL) references within web sites are maintained when changes occur in the locations where resources referenced by URLs are stored. A Referential Preservation Engine (RPE) maintains a database in which the location of web site documents and reference information are stored and updates various URL hyperlink references contained in the web pages on the site so that users can locate documents that have been moved to new storage locations. The RPE can also update links to external web sites by communicating with an RPE running on each external site. The RPE on the external site keeps track of the movement of linked documents on the sites and passes information pertaining to the new location of the linked documents to the local site, whereupon the links on the local web site pages are updated to reflect the new storage locations. The RPE also can track usage of a user's favorite sites and/or documents that are stored in an Internet browser and update the URL references for these favorites when the resources they are mapped to are moved (or renamed).

9 Claims, 13 Drawing figures

Full | Title | Citation | Front | Review | Classification | Date | Reference | | | Claims | KWIC | Draw. De

---

9. Document ID: US 6285999 B1

L6: Entry 9 of 11 File: USPT Sep 4, 2001

US-PAT-NO: 6285999
DOCUMENT-IDENTIFIER: US 6285999 B1

TITLE: Method for node ranking in a linked database

DATE-ISSUED: September 4, 2001

INVENTOR-INFORMATION:

NAME	CITY	STATE	ZIP CODE	COUNTRY

Page; Lawrence Stanford CA

ASSIGNEE-INFORMATION:

NAME	CITY	STATE	ZIP CODE	COUNTRY	TYPE CODE
The Board of Trustees of the Leland Stanford Junior University	Stanford	CA			02

APPL-NO: 09/004827 [PALM]
DATE FILED: January 9, 1998

PARENT-CASE:
CROSS-REFERENCES TO RELATED APPLICATIONS This application claims priority from U.S. provisional patent application Ser. No. 60/035,205 filed Jan. 10, 1997, which is incorporated herein by reference.

INT-CL: [07] G06 F 17/30

US-CL-ISSUED: 707/5; 707/7, 707/501
US-CL-CURRENT: 707/5; 707/7, 715/501.1

FIELD-OF-SEARCH: 707/100, 707/5, 707/7, 707/513, 707/1-3, 707/10, 707/104, 707/501, 345/440, 382/226, 382/229, 382/230, 382/231

PRIOR-ART-DISCLOSED:

U.S. PATENT DOCUMENTS

PAT-NO	ISSUE-DATE	PATENTEE-NAME	US-CL
4953106	August 1990	Gansner et al.	345/440
5450535	September 1995	North	395/140
5748954	May 1998	Mauldin	395/610
5752241	May 1998	Cohen	707/3
5832494	November 1998	Egger et al.	707/102
5848407	December 1998	Ishikawa et al.	707/2
6014678	January 2000	Inoue et al.	707/501

OTHER PUBLICATIONS

S. Jeromy Carriere et al, "Web Query: Searching and Visualizing the Web through Connectivity", Computer Networks and ISDN Systems 29 (1997). pp. 1257-1267.*
Wang et al "Prefetching in Worl Wide Web", IEEE 1996, pp. 28-32.*
Ramer et al "Similarity, Probability and Database Organisation: Extended Abstract", 1996, pp. 272.276.*
Craig Boyle "To link or not to link: An empirical comparison of Hypertext linking strategies". ACM 1992, pp. 221-231.*
L. Katz, "A new status index derived from sociometric analysis," 1953, Psychometricka, vol. 18, pp. 39-43.
C.H. Hubbell, "An input-output approach to clique identification sociometry," 1965, pp. 377-399.
Mizruchi et al., "Techniques for disaggregating centrality scores in social networks," 1996, Sociological Methodology, pp. 26-48.
E. Garfield, "Citation analysis as a tool in journal evaluation," 1972, Science, vol. 178, pp. 471-479.

Pinski et al., "Citation influence for journal aggregates of scientific publications: Theory, with application to the literature of physics," 1976, Inf. Proc. And Management, vol. 12, pp. 297-312.
N. Geller, "On the citation influence methodology of Pinski and Narin," 1978, Inf. Proc. And Management, vol. 14, pp. 93-95.
P. Doreian, "Measuring the relative standing of disciplinary journals," 1988, Inf. Proc. And Management, vol. 24, pp. 45-56.
P. Doreian, "A measure of standing for citation networks within a wider environment," 1994, Inf. Proc. And Management, vol. 30, pp. 21-31.
Botafogo et al., "Structural analysis of hypertext: Identifying hierarchies and useful metrics," 1992, ACM Trans. Inc. Systems, vol. 10, pp. 142-180.
Mark E. Frisse, "Searching for information in a hypertext medical handbook," 1988, Communications of the ACM, vol. 31, No. 7, pp. 880-886.
Massimo Marchiori, "The quest for correct information on the Web: Hyper search engines," 1997, Computer Networks and ISDN Systems, vol. 29, No. 8-13, pp. 1225-1235.
Oliver A. McBryan, "GENVL and WWWW: Tools for taming the web," 1994, Proceedings of the first International Wold Wide Web Conference, pp. 1-13.
Carriere et al., "WebQuery: Searching and visualizing the web through connectivity," 1997, Proc. 6.sup.th International World Wide Web Conference, pp. 1-14.
Arocena et al., "Applications of a web query language," 1997, Computer Networks and ISDN Systems, vol. 29, No. 8-13, pp. 1305-1316.
Jon M. Kleinberg, "Authoritative sources in a hyperlinked environment," 1998, Proc. Of the 9.sup.th Annual ACM-SIAM Symposium on Discrete Algorithms, pp. 668-677.
Henzinger et al., "Measuring index quality using random walks on the web", 1999, Proc. of the 8.sup.th International World Wide Web Conference, pp. 213-225.

ART-UNIT: 211

PRIMARY-EXAMINER: Black; Thomas

ASSISTANT-EXAMINER: Le; Uyen

ATTY-AGENT-FIRM: Harrity & Snyder L.L.P.

ABSTRACT:

A method assigns importance ranks to nodes in a linked database, such as any database of documents containing citations, the world wide web or any other hypermedia database. The rank assigned to a document is calculated from the ranks of documents citing it. In addition, the rank of a document is calculated from a constant representing the probability that a browser through the database will randomly jump to the document. The method is particularly useful in enhancing the performance of search engine results for hypermedia databases, such as the world wide web, whose documents have a large variation in quality.

29 Claims, 3 Drawing figures

☐ 10. Document ID: US 6138113 A

L6: Entry 10 of 11 File: USPT Oct 24, 2000

US-PAT-NO: 6138113
DOCUMENT-IDENTIFIER: US 6138113 A

TITLE: Method for identifying near duplicate pages in a hyperlinked database

DATE-ISSUED: October 24, 2000

INVENTOR-INFORMATION:

NAME	CITY	STATE	ZIP CODE	COUNTRY
Dean; Jeffrey	Menlo Park	CA		
Henzinger; Monika R.	Menlo Park	CA		

ASSIGNEE-INFORMATION:

NAME	CITY	STATE	ZIP CODE	COUNTRY	TYPE CODE
AltaVista Company	Palo Alto	CA			02

APPL-NO: 09/131469 [PALM]
DATE FILED: August 10, 1998

INT-CL: [07] G06 F 17/30

US-CL-ISSUED: 707/2; 707/100, 340/825.44
US-CL-CURRENT: 707/2; 707/100

FIELD-OF-SEARCH: 707/1-10, 707/100-104, 707/200-206, 709/226, 711/163, 340/825.22, 340/825.44, 370/390, 370/400, 370/408

PRIOR-ART-DISCLOSED:

U.S. PATENT DOCUMENTS

PAT-NO	ISSUE-DATE	PATENTEE-NAME	US-CL
5241305	August 1993	Fascenda et al.	340/825.44
5309433	May 1994	Cidon et al.	370/390
5335325	August 1994	Frank et al.	711/163
5345227	September 1994	Fascenda et al.	340/825.22
5425021	June 1995	Derby et al.	370/408
5483522	January 1996	Derby et al.	370/400
5917424	June 1999	Goldman et al.	340/825.44
5991809	November 1999	Kriegsman	709/226

ART-UNIT: 271

PRIMARY-EXAMINER: Ho; Ruay Lian

ATTY-AGENT-FIRM: Skjerven Morrill MacPherson LLP

ABSTRACT:

A method is described for identifying pages that are near duplicates in a linked

database. In the linked database, pages can have incoming links and outgoing links. Two pages are selected, a first page and a second page. For each selected page, the number of outgoing links is determined. The two pages are marked as near duplicates based on the number of common outgoing links for the two pages.

4 Claims, 2 Drawing figures

Full | Title | Citation | Front | Review | Classification | Date | Reference | | Claims | KWIC | Draw De

11. Document ID: US 5542024 A

L6: Entry 11 of 11 File: USPT Jul 30, 1996

US-PAT-NO: 5542024
DOCUMENT-IDENTIFIER: US 5542024 A

TITLE: Graphically used expert system tool background of the invention

DATE-ISSUED: July 30, 1996

INVENTOR-INFORMATION:

NAME	CITY	STATE	ZIP CODE	COUNTRY
Balint; George G.	Neshanic Station	NJ		
Hood; Douglas W.	Edison	NJ		

ASSIGNEE-INFORMATION:

NAME	CITY	STATE	ZIP CODE	COUNTRY	TYPE CODE
Johnson & Johnson		NJ			02

APPL-NO: 07/910814 [PALM]
DATE FILED: July 9, 1992

INT-CL: [06]